CLAIMS

What is claimed is:

1   1.   A method, comprising:

2   electronically capturing visual features associated with a speaker

3   speaking;

4   electronically capturing audio;

5   matching selective portions of the audio with the visual features; and

6   identifying the remaining portions of the audio as potential noise not

7   associated with the speaker speaking.

1   2.   The method of claim 1 further comprising:

2   electronically capturing additional visual features associated with a

3   different speaker speaking; and

4   matching some of the remaining portions of the audio from the

5   potential noise with the additional speaker speaking.

1   3.   The method of claim 1 further comprising generating parameters

2   associated with the matching and the identifying and providing the

3   parameters to a Bayesian Network which models the speaker speaking.

1   4.   The method of claim 1 wherein electronically capturing the visual

2   features further includes processing a neural network against electronic

3   video associated with the speaker speaking, wherein the neural network is

4   trained to detect and monitor a face of the speaker.

1   5.   The method of claim 4 further comprising filtering the detected face of

2   the speaker to detect movement or lack of movement in a mouth of the

3   speaker.

1   6.   The method of claim 1 wherein matching further includes comparing

2   portions of the captured visual features against portions of the captured

3   audio during a same time slice.

1   7.   The method of claim 1 further comprising suspending the capturing of

2   audio during periods where select ones of the captured visual features

3   indicate that the speaker is not speaking.

1   8.   A method, comprising:

2   monitoring an electronic video of a first speaker and a second

3   speaker;

4   concurrently capturing audio associated with the first and second

5   speaker speaking;

6   analyzing the video to detect when the first and second speakers are

7   moving their respective mouths; and

8   matching portions of the captured audio to the first speaker and other

9   portions to the second speaker based on the analysis.

1   9.   The method of claim 8 further comprising modeling the analysis for

2   subsequent interactions with the first and second speakers.

1   10.   The method of claim 8 wherein analyzing further includes processing

2   a neural network for detecting faces of the first and second speakers and

3   processing vector classifying algorithms to detect when the first and second

4   speakers' respective mouths are moving or not moving.

1   11.   The method of claim 8 further comprising separating the electronic

2   video from the concurrently captured audio in preparation for analyzing.

1   12.   The method of claim 8 further comprising suspending the capturing

2   of audio when the analysis does not detect the mouths moving for the first

3   and second speakers.

1   13.   The method of claim 8 further comprising identifying selective

2   portions of the captured audio as noise if the selective portions have not

3   been matched to the first speaker or the second speaker.

1   14.   The method of claim 8 wherein matching further includes identifying

2   time dependencies associated with when selective portions of the electronic

3   video were monitored and when selective portions of the audio were

4   captured.

1   15.   A system, comprising:

2   a camera;

3   a microphone; and

4   a processing device, wherein the camera captures video of a speaker

5   and communicates the video to the processing device, the microphone

6   captures audio associated with the speaker and an environment of the

7   speaker and communicates the audio to the processing device, the

8   processing device includes instructions that identifies visual features of the

9   video where the speaker is speaking and uses time dependencies to match

10   portions of the audio to those visual features.

1   16.   The system of claim 15 wherein the captured video also includes

2   images of a second speaker and the audio includes sounds associated with

3   the second speaker, and wherein the instructions matches some portions of

4   the audio to the second speaker when some of the visual features indicate

5   the second speaker is speaking.

1   17.   The system of claim 15 wherein the instructions interact with a neural

2   network to detect a face of the speaker from the captured video.

1   18.   The system of claim 17 wherein the instructions interact with a pixel

2   vector algorithm to detect when a mouth associated with the face moves or

3   does not move within the captured video.

1   19.   The system of claim 18 wherein the instructions generate parameter

2   data that configures a Bayesian network which models subsequent

3   interactions with the speaker to determine when the speaker is speaking

4   and to determine appropriate audio to associate with the speaker speaking

5   in the subsequent interactions.

1   20.   A machine accessible medium having associated instructions, which

2   when accessed, results in a machine performing:

3   separating audio and video associated with a speaker speaking;

4   identifying visual features from the video that indicate a mouth of the

5   speaker is moving or not moving; and

6   associating portions of the audio with selective ones of the visual

7   features that indicate the mouth is moving.

1   21.   The medium of claim 20 further including instructions for associating

2   other portions of the audio with different ones of the visual features that

3   indicate the mouth is not moving.

1   22.   The medium of claim 20 further including instructions for:

2   identifying second visual features from the video that indicate a

3   different mouth of another speaker is moving or not moving; and

4   associating different portions of the audio with selective ones of the

5   second visual features that indicate the different mouth is moving.

1   23.   The medium of claim 20 wherein the instructions for identifying further

2   include instructions for:

3   processing a neural network to detect a face of the speaker; and

4   processing a vector matching algorithm to detect movements of the

5   mouth of the speaker within the detected face.

1   24.   The medium of claim 20 wherein the instructions for associating

2   further include instructions for matching same time slices associated with a

3   time that the portions of the audio were captured and the same time during

4   which the selective ones of the visual features were captured within the

5   video.

1   25.   An apparatus, residing in a computer-accessible medium, comprising:

2   face detection logic;

3   mouth detection logic; and

4   audio-video matching logic, wherein the face detection logic detects a

5   face of a speaker within a video, the mouth detection logic detects and

6   monitors movement and non-movement of a mouth included within the face

7   of the video, and the audio-video matching logic matches portions of

8   captured audio with any movements identified by the mouth detection logic.

1   26.   The apparatus of claim 25 wherein the apparatus is used to configure

2   a Bayesian network which models the speaker speaking.

1   27.   The apparatus of claim 25 wherein the face detection logic comprises

2   a neural network.

1   28.   The apparatus of claim 25 wherein the apparatus resides on a

2   processing device and the processing device is interfaced to a camera and

3   a microphone.